



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 138997**

**TO: Cynthia Wilder**  
**Location: REM/2A35/2C18**  
**Wednesday, April 20, 2005**  
**Art Unit: 1637**

**Case Serial Number: 09/529397**

**From: David Schreiber**  
**Location: Biotech-Chem Library**  
**Remsen E01A61**  
**Phone: 571-272-2526**

**David.Schreiber@uspto.gov**

### **Search Notes**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 19, 2005, 13:28:09 ; Search time 427 Seconds  
(without alignments)  
1386.356 Million cell updates/sec

Title: US-09-529-397C-25  
Perfect score: 100  
Sequence: 1 999gagagagaaucaucg.....uagcagcagcagcagcagc 100

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	100	100.0	100	3	AAZ99048	RNA aptam
2	100	100.0	100	12	ADJ10063	Synthetic
3	100	100.0	100	12	ADJ10064	Synthetic
4	100	100.0	200	12	ADJ10065	Synthetic
5	90	90.0	90	3	AAZ99050	RNA aptam
6	84.4	84.4	98	12	ADJ10053	RNA aptam
7	80	80.0	80	3	AAZ99051	RNA aptam
8	50.6	50.6	147	12	ADJ10066	Synthetic
9	45	45.0	45	3	AAZ99076	Ras target
10	30.6	30.6	748	3	AAZ75113	Human ORF
11	30.6	30.6	748	6	ABN24508	Human ORF
12	30.6	30.6	15776	8	AAZ59583	Propionib
13	30.6	30.6	15776	8	ACF64512	Propionib
14	30.4	30.4	366	9	ADA29321	DNA encod
15	30	30.0	98	3	AAZ99049	RNA aptam
16	30	30.0	90616	12	ADQ97596	Mouse can
17	29.8	29.8	4192	3	AAZ57942	Poplar fl
18	29.8	29.8	4285	4	AAZ85391	Nucleotid
19	29.8	29.8	4285	6	ABK88484	Poplar pr
20	29.8	29.8	4285	9	ACA62517	Poplar ho

21	29	29.0	29	3	AAZ99084	Ras target
22	29	29.0	29	12	ADJ10055	Synthetic
23	28.8	28.8	675	8	ACA52761	Prokaryot
24	28.6	28.6	2000	6	ABZ15103	Arabidops
25	28.2	28.2	1161	5	ABV24321	Human pro
26	28.2	28.2	1665	2	AAZ58401	Vicia sat
27	28	28.0	29	12	ADJ10058	PCR prime
28	28	28.0	29	12	ADJ10057	PCR prime
29	28	28.0	29	12	ADJ10056	PCR prime
30	28	28.0	110000	6	ABX08336	Continuation (2 of
31	28	28.0	110000	12	ADJ25985	Continuation (2 of
32	28	28.0	110000	12	ADN97989	Continuation (2 of
33	28	28.0	110000	12	ADO50281	Continuation (2 of
34	27.8	27.8	1093	13	ADS47230	Bacterial
35	27.8	27.8	99291	13	ABD33549	Human can
36	27.6	27.6	732	4	AAI94947	Human neu
37	27.6	27.6	732	8	ABT42823	Human neu
38	27.4	27.4	3791	2	AAZ63575	Chicken b
39	27	27.0	27	3	AAZ99087	Ras target
40	27	27.0	29	12	ADJ10059	PCR prime
41	27	27.0	208	3	AAZ17425	Human sec
42	27	27.0	522	3	AAZ69206	Bacteriop
43	27	27.0	56506	3	AAZ69168	Bacteriop
44	26.8	26.8	415	3	AAZ21986	Human sec
45	26.8	26.8	1116	5	AAZ51835	HIV prote

ALIGNMENTS

RESULT 1  
AAZ99048  
ID AAZ99048 standard; RNA; 100 BP.  
XX  
AC AAZ99048;  
XX  
DT 21-JUN-2000 (first entry)  
DE  
DE RNA aptamer #25 for binding Ras target protein.  
XX  
KW Ras target protein; malignant tumour; signal transduction regulation;  
KW cell proliferation; cell differentiation; aptamer; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
PN HQ200009684-A1.  
XX  
PD 24-FEB-2000.  
XX  
PF 13-AUG-1999; 99WO-JP004399.  
XX  
PR 14-AUG-1998; 98JP-00242596.  
PR 24-NOV-1998; 98JP-00333284.  
XX  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX Yokoyama S, Hirao I, Sakamoto K;  
XX WPI; 2000-224330/19.  
XX  
XX Nucleic acid e.g. RNA aptamer capable of binding specifically to Ras  
XX target protein like Raf-1, useful in drug compositions to treat and  
XX diagnose malignant tumors and inflammation.  
XX  
XX Claim 6; Page 41; 59pp; Japanese.  
XX  
XX The invention relates to novel nucleic acids which bind specifically to  
XX the target protein of Ras, e.g. Raf-1. RNA aptamer (AAZ99024-299051)  
XX based on these sequences are useful in the treatment and diagnosis of  
XX malignant tumours and inflammation. The nucleic acids can be used to  
XX formulate medicinal compositions that are useful in the treatment of  
XX malignant tumours and inflammation and for disease diagnosis by binding  
XX specifically to Ras target protein and regulating transmission of signal

11/24/98

(PD) Priority document

CC causing proliferation or differentiation of cells  
XX  
SQ Sequence 100 BP; 25 A; 22 C; 24 G; 0 T; 29 U; 0 Other;  
Query Match 100.0%; Score 100; DB 3; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.7e-25;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGAGUGGAGGAUUAUCGAGGCAUAUGGACUCCGUCUCCUUAACAGUUUA 60  
Db 1 GGGAGUGGAGGAUUAUCGAGGCAUAUGGACUCCGUCUCCUUAACAGUUUA 60  
QY 61 AUUGGUUUUAGCAUAGCCUUAAGCGACAGCAAGCUUCUCG 100  
Db 61 AUUGGUUUUAGCAUAGCCUUAAGCGACAGCAAGCUUCUCG 100  
RESULT 2  
ADJ10063  
ID ADJ10063 standard; DNA; 100 BP.  
XX  
AC ADJ10063;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Synthetic RNA 9A (100-mer).  
XX  
KW ss; unnatural base; 5'-substituted-2-oxo(1H)-pyridin-3-yl;  
KW photoactive group; biotin; fluorescent molecule;  
KW 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.  
XX  
OS Synthetic.  
XX  
PN WO2004007713-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 28-FEB-2003; 2003WO-JP002342.  
XX  
PR 17-JUL-2002; 2002JP-00208568.  
XX  
PA (RIKE ) RIKEN KK.  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
PI Hirao I, Yokoyama S, Hirao M, Mitsui T;  
XX  
DR WPI; 2004-122944/12.  
XX  
PT Nucleosides or nucleotides comprising novel base particularly 5-substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing functional nucleic acids as reagents or therapeutics in treating diseases.  
XX  
PS Disclosure; Fig 10; 78pp; Japanese.  
XX  
CC This invention relates to nucleosides or nucleotides that a novel unnatural base, namely the 5'-substituted-2-oxo(1H)-pyridin-3-yl group as a base. Specifically, it refers to nucleic acid molecules integrated with a base substituted at the 5-position with either a photoactive group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group, biotin or a derivative thereof or a fluorescent molecule selected from fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrhodamine or derivatives thereof. The present invention describes preparing nucleic acids by carrying out transcription, replication or reverse transcription with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as

CC causing proliferation or differentiation of cells  
XX  
SQ Sequence 100 BP; 25 A; 22 C; 24 G; 0 T; 29 U; 0 Other;  
Query Match 100.0%; Score 100; DB 12; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.7e-25;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGAGUGGAGGAUUAUCGAGGCAUAUGGACUCCGUCUCCUUAACAGUUUA 60  
Db 1 GGGAGUGGAGGAUUAUCGAGGCAUAUGGACUCCGUCUCCUUAACAGUUUA 60  
QY 61 AUUGGUUUUAGCAUAGCCUUAAGCGACAGCAAGCUUCUCG 100  
Db 61 AUUGGUUUUAGCAUAGCCUUAAGCGACAGCAAGCUUCUCG 100  
RESULT 3  
ADJ10064  
ID ADJ10064 standard; DNA; 100 BP.  
XX  
AC ADJ10064;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Synthetic RNA 9A(51y87) aptamer (100-mer).  
XX  
KW ss; unnatural base; 5'-substituted-2-oxo(1H)-pyridin-3-yl;  
KW photoactive group; biotin; fluorescent molecule;  
KW 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.  
XX  
OS Synthetic.  
XX  
FH Key modified\_base 87  
FT /tag= a  
FT /mod\_base= OTHER  
FT /note= "OTHER= 51y"  
XX  
PN WO2004007713-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 28-FEB-2003; 2003WO-JP002342.  
XX  
PR 17-JUL-2002; 2002JP-00208568.  
XX  
PA (RIKE ) RIKEN KK.  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
PI Hirao I, Yokoyama S, Hirao M, Mitsui T;  
XX  
DR WPI; 2004-122944/12.  
XX  
PT Nucleosides or nucleotides comprising novel base particularly 5-substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing functional nucleic acids as reagents or therapeutics in treating diseases.  
XX  
PS Disclosure; Fig 10; 78pp; Japanese.  
XX  
CC This invention relates to nucleosides or nucleotides that a novel unnatural base, namely the 5'-substituted-2-oxo(1H)-pyridin-3-yl group as a base. Specifically, it refers to nucleic acid molecules integrated with a base substituted at the 5-position with either a photoactive group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group, biotin or a derivative thereof or a fluorescent molecule selected from fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrhodamine or derivatives thereof. The present invention describes preparing nucleic acids by carrying out transcription, replication or reverse transcription with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as



```
Query Match      90.0%; Score 90; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 9e-22;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGUGGAGGAUAUUCGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 60
   |||||
Db 1 GGGAGUGGAGGAUAUUCGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 60
   |||||

QY 61 AUUGGUUUUAGCAUAGCCUUAAGCGACGC 90
   |||||
Db 61 AUUGGUUUUAGCAUAGCCUUAAGCGACGC 90
   |||||

RESULT 6
ADJ10053
ID ADJ10053 standard; RNA; 98 BP.
AC ADJ10053;
DT 17-JUN-2004 (first entry)
XX RNA aptamer (RNA9A) SeqID 1.
DE ss; unnatural base; 5'-substituted-2-oxo(1H)-pyridin-3-yl;
KW photoreactive group; biotin; fluorescent molecule;
KW 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.
XX Unidentified.
OS WO2004067713-A1.
PN 32-JAN-2004.
PD 28-FEB-2003; 2003WO-JP002342.
XX 17-JUL-2002; 2002JP-00208568.
PR (RIKE) RIKEN KK.
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
PI Hirao I, Yokoyama S, Hirao M, Mitsui T;
XX WPI; 2004-122944/12.
DR Nucleosides or nucleotides comprising novel base particularly 5-
PT substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing
PT functional nucleic acids as reagents or therapeutics in treating
PI diseases.
PS Example 4; SEQ ID NO 1; 78pp; Japanese.
XX This invention relates to nucleosides or nucleotides that a novel
CC unnatural base, namely the 5'-substituted-2-oxo(1H)-pyridin-3-yl group
CC as a base. Specifically, it refers to nucleic acid molecules integrated
CC with a base substituted at the 5-position with either a photoreactive
CC group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group,
CC biotin or a derivative thereof or a fluorescent molecule selected from
CC fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrodamine or
CC derivatives thereof. The present invention describes preparing nucleic
CC acids by carrying out transcription, replication or reverse transcription
CC with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as
CC base which is applied as template for the integration of any of the
CC nucleotides into its complementary position. Accordingly, the nucleosides
CC and nucleotides are useful in developing functional nucleic acids
CC including antisense DNAs and RNAs, ribozymes and aptamers as reagents or
CC therapeutics in treating diseases by gene therapy. The produced nucleic
CC acids are chemically stable with improved ease of amplification.
CC replication and transcription for their preparation. This polynucleotide
CC sequence is an RNA aptamer (RNA9A) given in an exemplification of the
XX invention.
SQ Sequence 98 BP; 24 A; 21 C; 24 G; 0 T; 29 U; 0 Other;

Query Match      84.4%; Score 84.4; DB 12; Length 98;
Best Local Similarity 97.0%; Pred. No. 8.9e-20;
Matches 97; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1 GGGAGUGGAGGAUAUUCGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 60
   |||||
Db 1 GGGAGUGGAGGAUAUUC-UGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 58
   |||||

QY 61 AUUGGUUUUAGCAUAGCCUUAAGCGACGACGAGCUUUGC 100
   |||||
Db 59 AUUGGUUUUAGCAUAGCCUUAAGCGACGACGACGAGCUUUGC 98
   |||||

RESULT 7
AAZ99051
ID AAZ99051 standard; RNA; 80 BP.
XX AAZ99051;
DT 21-JUN-2000 (first entry)
XX RNA aptamer #28 for binding Ras target protein.
DE Ras target protein; malignant tumour; signal transduction regulation;
KW cell proliferation; cell differentiation; aptamer; inflammation; ss.
XX Homo sapiens.
OS WO200009684-A1.
PN 24-FEB-2000.
PD 13-AUG-1999; 99WO-JP004399.
XX 14-AUG-1998; 98JP-00242596.
PR 24-NOV-1998; 98JP-00333284.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA Yokoyama S, Hirao I, Sakamoto K;
PI WPI; 2000-224330/19.
DR Nucleic acid e.g. RNA aptamer capable of binding specifically to Ras
PT target protein like Raf-1, useful in drug compositions to treat and
PT diagnose malignant tumors and inflammation.
XX Claim 6; Page 42; 59pp; Japanese.
XX The invention relates to novel nucleic acids which bind specifically to
CC the target protein of Ras, e.g. Raf-1. RNA aptamer (AAZ99051)
CC based on these sequences are useful in the treatment and diagnosis of
CC malignant tumors and inflammation. The nucleic acids can be used to
CC formulate medicinal compositions that are useful in the treatment of
CC malignant tumors and inflammation and for disease diagnosis by binding
CC specifically to Ras target protein and regulating transmission of signal
CC causing proliferation or differentiation of cells
XX
SQ Sequence 80 BP; 20 A; 16 C; 19 G; 0 T; 25 U; 0 Other;

Query Match      80.0%; Score 80; DB 3; Length 80;
Best Local Similarity 100.0%; Pred. No. 3e-18;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGUGGAGGAUAUUCGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 60
   |||||
Db 1 GGGAGUGGAGGAUAUUCGAGGCAUAUGUGGACUCCGUCUCCUUAACAGUUAUA 60
   |||||

QY 61 AUUGGUUUUAGCAUAGCCU 80
   |||||
Db 61 AUUGGUUUUAGCAUAGCCU 80
   |||||
```







```
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

QY 18 UCGAGGCAUAGUCGACUCCGUCUCCUUAACACGAGUUAUAAUUGGUUUUAGCAUAG 77
Db 317 TTGCGGTGACCTCGACGCGCAGCGTCTTTCAGACCGGTTCATGAAACGGTTGTAGTGCATG 258
QY 78 CCUUAAGCGACGACGCAAGC 94
Db 257 CCTGGGCGACGGGAAGC 241

RESULT 12
ID AAS59583
XX AC AAS59583;
XX DT 13-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein encoding DNA #78.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant; ds.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US012865.
XX PR 21-APR-2000; 2000US-0199047P.
XX PR 02-JUN-2000; 2000US-0208841P.
XX PR 07-JUL-2000; 2000US-0216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris.
XX PS Claim 1; SEQ ID NO 78; 1069pp; English.
XX CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
XX CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX CC associated DNA sequences are used in the treatment, prevention and
XX CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
XX CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
XX CC particularly involved in the inflammatory lesions associated with acne
XX CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX CC patient comprises contacting a sample with a binding agent that binds to
XX CC the proteins of the invention and determining the amount of bound protein
XX CC in the sample. The polypeptides may be used as antigens in the production
XX CC of antibodies specific for P. acnes proteins. These antibodies can be
XX CC used to downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX CC polypeptides shown in AAU57347-AAU57508 and AAU67590-AAU67591. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
```

```
SQ Sequence 15776 BP; 3393 A; 4931 C; 4483 G; 2969 T; 0 U; 0 Other;

Query Match 30.6%; Score 30.6; DB 4; Length 15776;
Best Local Similarity 46.8%; Pred. No. 4.5;
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

QY 18 UCGAGGCAUAGUCGACUCCGUCUCCUUAACACGAGUUAUAAUUGGUUUUAGCAUAG 77
Db 8427 TTGCGGTGACCTCGACGCGCAGCGTCTTTCAGACCGGTTCATGAAACGGTTGTAGTGCATG 8486
QY 78 CCUUAAGCGACGACGCAAGC 94
Db 8487 CCTGGGCGACGGGAAGC 8503

RESULT 13
ACF64512
ID ACF64512 standard; DNA; 15776 BP.
XX AC ACF64512;
XX DT 17-OCT-2003 (first entry)
XX DE Propionibacterium acnes DNA contig sequence #78.
XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
XX KW immunostimulant; immune response; vaccine; ds.
XX OS Propionibacterium acnes.
XX PN WO2003033515-A1.
XX PD 24-APR-2003.
XX PF 11-OCT-2002; 2002WO-US032727.
XX PR 15-OCT-2001; 2001US-00978825.
XX PA (CORI-) CORIXA CORP.
XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
XX PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
XX PI Barth B, Vallieve-Doughlass J;
XX WPI; 2003-381789/36.
XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
XX PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
XX PT or for stimulating an immune response specific for a P. acnes protein.
XX PS Claim 1; SEQ ID NO 78; 1481pp; English.
XX CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
XX CC encoding a Propionibacterium acnes protein. The invention also relates to
XX CC polypeptides encoded by the polynucleotides (ABW35624-ABW64536) and to
XX CC immunogenic fragments of P. acnes polypeptides. The invention
XX CC additionally encompasses expression vectors and host cells comprising a
XX CC polynucleotide of the invention; antibodies against polypeptides of the
XX CC invention; fusion proteins comprising a polypeptide of the invention; a
XX CC method for stimulating an immune response specific for a P. acnes
XX CC polypeptide and an isolated T cell population comprising T cells prepared
XX CC via this method; a vaccine composition (comprising P. acnes polypeptides,
XX CC polynucleotides, antibodies, fusion proteins, T cell populations, or
XX CC antigen-presenting cells that express the polypeptide); a method and kit
XX CC for detecting or determining the presence or absence of P. acnes in a
XX CC patient; and a method for inhibiting the development of P. acnes in a
XX CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
XX CC proteins, T cell populations or antigen-presenting cells that express the
XX CC polypeptides are useful for diagnosing, preventing or treating acne
XX CC vulgaris, or for stimulating an immune response specific for a P. acnes
XX CC protein. The polynucleotides can also be used as probes or primers for
XX CC nucleic acid hybridisation. The vaccine composition is useful for the
XX CC stimulation of an immune response against P. acnes, or for treating acne,
```

CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a P. acnes DNA contig which is specifically claimed  
CC in the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 15776 BP; 3393 A; 4931 C; 4483 G; 2969 T; 0 U; 0 Other;

Query Match 30.6%; Score 30.6; DB 8; Length 15776;

Best Local Similarity 46.8%; Pred. No. 4.5; Mismatches 29; Indels 0; Gaps 0;

Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

OY 18 UCGAGCAUAUGUGGACUCGCUUCCUUAACAGUUAUAAUUGGUUUAUGCAUUG 77

Db 8427 TTGGGTGACCTCGACGCGCGGTCTTCACACCGTTCATGACGCGTTGAGTCATG 8486

OY 78 CCUAGCGACAGCAGC 94

Db 8487 CCTGGCAGCGGAAGC 8503

RESULT 14

ADA29321/c

ID ADA29321 standard; DNA; 366 BP.

AC ADA29321;

XX

XX

DT 20-NOV-2003 (first entry)

XX

DE DNA encoding Acinetobacter baumannii protein #608.

XX

KW ds; Gene; Acinetobacter baumannii; bacterial disease; antibacterial;

XX

KW vaccine; plant biocontrol agent.

XX

OS Acinetobacter baumannii.

XX

PN US6562958-B1.

XX

PD 13-MAY-2003.

XX

PF 04-JUN-1999; 99US-00328352.

XX

PR 09-JUN-1998; 98US-0088701P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Breton G, Bush D;

XX

DR WPI; 2003-576092/54.

XX

DR P-PSDB; ADA33447.

XX

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.

XX

PS Example; SEQ ID NO 608; 328pp; English.

XX

CC The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX

CC The A. baumannii nucleic acids and polypeptides are useful as reagents

XX

CC for diagnosing a bacterial disease, as components of antibacterial

XX

CC vaccines, as targets for antibacterial drugs, to detect the presence of

XX

CC A. baumannii and other Acinetobacter species in a sample, in screening

XX

CC compounds for the ability to interfere with the A. baumannii life cycle

XX

CC or to inhibit A. baumannii infection, and as biocontrol agents for

XX

CC plants. The present sequence represents DNA encoding an A. baumannii

XX

CC protein.

XX

SQ Sequence 366 BP; 131 A; 65 C; 63 G; 107 T; 0 U; 0 Other;

Query Match

Best Local Similarity 30.4%; Score 30.4; DB 9; Length 366;

Matches 35; Conservative 17; Mismatches 36; Indels 0; Gaps 0;

OY 1 GGGAGUGGAGGAAUUAUCGAGGCAUAUGUGCAUCUCGCUUCCUUAACCAAGUUAUA 60  
Db 101 GGGTGTATGATATAGCTTTCTATATGTGCACAGTAATTGTGCAACAGTANTAG 42  
OY 61 AUUGUUUUAUGCAUAUGCCUUAUGCGACA 88  
Db 41 TCTGTTTATGCTATATTTCTAGATTCA 14

RESULT 15

AAZ99049

ID AAZ99049 standard; RNA; 98 BP.

XX

AC AAZ99049;

XX

DT 21-JUN-2000 (first entry)

XX

DE RNA aptamer #26 for binding Ras target protein.

XX

KW Ras target protein; malignant tumour; signal transduction regulation;

XX

KW cell proliferation; cell differentiation; aptamer; inflammation; ss.

XX

OS Homo sapiens.

XX

PN WO200009684-A1.

XX

PD 24-FEB-2000.

XX

PF 13-AUG-1999; 99WO-JP004399.

XX

PR 14-AUG-1998; 98JP-00242596.

XX

PR 24-NOV-1998; 98JP-00333284.

XX

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX

PI Yokoyama S, Hirao I, Sakamoto K;

XX

DR WPI; 2000-224330/19.

XX

PT Nucleic acid e.g. RNA aptamer capable of binding specifically to Ras

XX

PT target protein like Raf-1, useful in drug compositions to treat and

XX

PT diagnose malignant tumors and inflammation.

XX

PS Claim 6; Page 41; 59pp; Japanese.

XX

CC The invention relates to novel nucleic acids which bind specifically to  
CC the target protein of Ras, e.g. Raf-1. RNA aptamer (AAZ99024-299051)  
CC based on these sequences are useful in the treatment and diagnosis of  
CC malignant tumours and inflammation. The nucleic acids can be used to  
CC formulate medicinal compositions that are useful in the treatment of  
CC malignant tumours and inflammation and for disease diagnosis by binding  
CC specifically to Ras target protein and regulating transmission of signal  
CC causing proliferation or differentiation of cells

XX

SQ Sequence 98 BP; 24 A; 24 C; 26 G; 0 T; 24 U; 0 Other;

Query Match

Best Local Similarity 30.0%; Score 30; DB 3; Length 98;

Matches 63; Conservative 0; Mismatches 35; Indels 2; Gaps 1;

OY 1 GGGAGUGGAGGAAUUAUCGAGGCAUAUGUGCAUCUCGCUUCCUUAACCAAGUUAUA 60

Db 1 GGGAGUGGAGGAAUUAUCGAGGCAUAUGUGCAUCUCGCGGAGGAGGUAUUAUUAU 60

OY 61 AUUGUUUUAUGCAUAUGCCUUAUGCGACAAGCUUCUGC 100

Db 61 CCUACACUU--CUAUGCCUUAUGCGACAGCAAGCUUCUGC 98

Search completed: April 19, 2005, 16:27:31

Job time : 432 secs





```
RESULT 5
US-10-425-114-30801/c
; Sequence 30801, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 28
; SEQ ID NO 30801
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73080F09_FLI
US-10-425-114-30801

Query Match      28.2%; Score 28.2; DB 17; Length 586;
Best Local Similarity 37.1%; Pred. No. 19;
Matches 33; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 8 GAGGAUAUUCGAGGCAUAUGUCGACUCCGUCUUAACCAACCAAGUUAUAAUUGGUU 67
DB 92 GGGGCATTTTATGTTGTTATTAGGATTCCTCTGCTCCACCGCATCAATACTCGGTT 33
QY 68 UUAGCAUAUGCCUUAAGCGACAGCAAGUU 96
DB 32 CCAGCAAGCGCATTTAGGGAAGGAAGATT 4

RESULT 6
US-10-357-930-24310/c
; Sequence 24310, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Endege, wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; FILE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24310
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-24310

Query Match      28.2%; Score 28.2; DB 18; Length 1161;
Best Local Similarity 37.1%; Pred. No. 24;
Matches 33; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 6 UGGAGGAUAUUCGAGGCAUAUGUCGACUCCGUCUUAACCAACCAAGUUAUAAUUGGU 65
DB 878 TGAAATAGCCTTAATGCAATTTCTTCACAGTAATTCAGCCAAATAATTTTAAATCAG 819
QY 66 UUUUAGCAUAUGCCUUAAGCGACAGCAAGC 94
DB 818 GTTAACTTATGCTTTAAAGAAATTAAGC 790

RESULT 7
US-10-767-701-13576/c
; Sequence 13576, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13576
; LENGTH: 1612
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS1303_1
US-10-767-701-13576

Query Match      28.2%; Score 28.2; DB 18; Length 1612;
Best Local Similarity 38.2%; Pred. No. 27;
Matches 34; Conservative 17; Mismatches 38; Indels 0; Gaps 0;

QY 8 GAGGAUAUUCGAGGCAUAUGUCGACUCCGUCUUAACCAACCAAGUUAUAAUUGGUU 67
DB 1152 GGGGCATTTTGTGTTGTTATTAGGATTCCTCTGCTCCACTGCGTCACAACTCGGTT 1093
QY 68 UUAGCAUAUGCCUUAAGCGACAGCAAGUU 96
DB 1092 CCAGCAAGCGCATTTAGGGAAGGAAGATT 1064

RESULT 8
US-10-425-115-121623/c
; Sequence 121623, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 121623
; LENGTH: 1851
; TYPE: DNA
; ORGANISM: Zea mays
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_42398C.1
US-10-425-115-121623

Query Match      28.2%; Score 28.2; DB 18; Length 1851;
Best Local Similarity 37.1%; Pred. No. 29;
Matches 33; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 8 GAGGAUAUCAGGCAUAUGGACUCCGUCUCCUCAAACAGUUAUAAUUGGUU 67
DB 1333 GGGCATTTTATGTTGTTATTAGGATTCCTGCTCCAGCCGATCATAACTCGGTT 1274

QY 68 UUAGCAUAUGCCUUCAGGACAGCAAGCUU 96
DB 1273 CCAGCAGCGCATTAGGGAAGAGATTT 1245

RESULT 9
US-10-067-514-1/c
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretaarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reyndisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match      28.0%; Score 28; DB 14; Length 1691139;
Best Local Similarity 41.7%; Pred. No. 4e+02;
Matches 25; Conservative 15; Mismatches 20; Indels 0; Gaps 0;

QY 12 AAUUAUCAGGCAUAUGGACUCCGUCUCCUCAAACAGUUAUAAUUGGUUAG 71
DB 127542 AGTTGATCATGCTTTTGGAGGAGCTTATCCCTTCAAAGCAGTAATAAAAGTGTTTAG 127483

RESULT 10
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
; APPLICANT: Gretaarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reyndisdottir, Sigridur Th.
; APPLICANT: Thorleifsson, Gudmar
; APPLICANT: Gulcher, Jeffrey
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; FILE REFERENCE: 2345.2010-005
; CURRENT APPLICATION NUMBER: US/10/419,723
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/255;120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-419-723-1

Query Match      28.0%; Score 28; DB 17; Length 1691139;
Best Local Similarity 41.7%; Pred. No. 4e+02;
Matches 25; Conservative 15; Mismatches 20; Indels 0; Gaps 0;

QY 12 AAUUAUCAGGCAUAUGGACUCCGUCUCCUCAAACAGUUAUAAUUGGUUAG 71
DB 127542 AGTTGATCATGCTTTTGGAGGAGCTTATCCCTTCAAAGCAGTAATAAAAGTGTTTAG 127483

RESULT 11
US-10-369-493-25760
; Sequence 25760, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25760
; LENGTH: 1093
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-25760

Query Match      27.8%; Score 27.8; DB 17; Length 1093;
Best Local Similarity 38.0%; Pred. No. 33;
Matches 27; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 4 AGUGGAGGAUAUUAUCGAGGCAUAUGGACUCCGUCUCCUCAAACAGUUAUAAU 63
DB 213 AGTGGACTCTTTCATAGACAATGAAGCAGGATCCGTTTACCTTATACGAGTTTAAATC 272
QY 64 GGUUUUAGCAU 74
DB 273 GTTTGTTCTAT 283

RESULT 12
US-10-322-281-744/c
; Sequence 744, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 744
; LENGTH: 99291
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-322-281-744

Query Match      27.8%; Score 27.8; DB 18; Length 99291;
Best Local Similarity 31.0%; Pred. No. 1.7e+02;
Matches 22; Conservative 22; Mismatches 27; Indels 0; Gaps 0;
```

Qy	12	AATUUCACUGGAGCAUAUGCACUCGCUCCUCAAACACGAGUUAAAUGGUUUAG	71
		:   :   :   :   :   :   :   :   :   :   :   :	
Dd	23527	AATAAATTGACACTCTCAGATACCTTTTGGTTGCACAATATAAAATGGTTTTAC	23468
Qy	72	CAUAUGCCUUUA	82
		:   :   :   :   :	
Dd	23467	AATTTGGTTTA	23457

RESULT 13  
US-10-479-081, Application US-10479081  
; Sequence 104, Application US/10479081  
; Publication No. US20050059001A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWARA, AKIRA  
; TITLE OF INVENTION: NUCLEIC ACIDS ISOLATED FROM NEUROBLASTOMA  
; FILE REFERENCE: 7388-80893  
; CURRENT APPLICATION NUMBER: US/10/479,081  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: PCT/JP02/05295  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-163666  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: JP 2001-255260  
; PRIOR FILING DATE: 2001-08-24  
; NUMBER OF SEQ ID NOS: 742  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 104  
; LENGTH: 732  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2)..(2)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (15)..(15)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (73)..(73)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (87)..(87)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (90)..(90)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (104)..(105)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (566)..(566)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (642)..(642)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (650)..(650)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature

```

; LOCATION: (706)..(706)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (711)..(711)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (732)..(732)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; OTHER INFORMATION: mbla-03439-f
; US-10-479-081-104

Query Match          27.6%; Score 27.6; DB 19; Length 732;
Best Local Similarity 47.6%; Pred. No. 33;
Matches 39; Conservative 9; Mismatches 34; Indels 0; Gaps 0

Qy 10 GGAUUAUCGAGGCAUAUGCGACUCGCGUCUCCUCAAACACGUAUAUAAUUGGUUU 69
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 161 GGAACATCATCCAGACTATCGACACATTTGCTCTTCTCCAGGCCACAGTAAGCAAGAAGA 220
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Qy 70 AGCAUAUGCCUUAAGCGACAGCA 91
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 221 TGCAGAAGACTTTAGCTGCAGCA 242
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT 14  
 US-10-097-111-47  
 ; Sequence 47, Application US/10097111  
 ; Publication No. US20030138771A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PELLETIER, JERRY  
 ; APPLICANT: GROS, PHILLIPE  
 ; APPLICANT: DUBOW, MICHAEL  
 ; TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EPI THAT  
 ; FILE REFERENCE: 073406-0603  
 ; CURRENT APPLICATION NUMBER: US/10/097,111  
 ; CURRENT FILING DATE: 2002-07-24  
 ; PRIOR APPLICATION NUMBER: 05/676,412  
 ; PRIOR FILING DATE: 2000-09-29  
 ; PRIOR APPLICATION NUMBER: 60/157,218  
 ; PRIOR FILING DATE: 1999-09-30  
 ; NUMBER OF SEQ ID NOS: 552  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 47  
 ; LENGTH: 522  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-097-111-47

	Query Match	27.0%;	Score 27;	DB 15;	Length 522;
	Best Local Similarity	40.0%;	Pred. No. 47;	Mismatches	Gaps 0;
	Matches 30;	Conservative 15;	Mismatches 30;	Indels 0;	Gaps 0;
Qy	13	AUUCAGCGGCAUAUGCAGTCGCCGUCCUUCAAACAGUUAUAUAUUGUUUAGC	72		
Dd	186	ATTCTATGCAGACTTGACCACGCTGTCTTCTTTCAAGGAATGAACAATCGCTTTAGC	245		
Qy	73	AUAUGC CUUAGCGAC	87		
Dd	246	AAATGCAGTTGCAC	260		

RESULT 15  
US-10-097-111-10  
; Sequence 10, Application US/10097111  
; Publication No. US2003013877A1  
; GENERAL INFORMATION:  
; APPLICANT: PELLETIER, JERRY  
; APPLICANT: GROS, PHILLIPPE  
; APPLICANT: DUBOW, MICHAEL





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2005, 13:28:09 ; Search time 3095 Seconds  
(without alignments)  
1229.863 Million cell updates/sec

Title: US-09-529-397C-25  
Perfect score: 100  
Sequence: 1 999agugaggaaucaucg.....uagcgacgacgaagcuucgc 100

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hc:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gss1:  
9: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.6	34.6	642	2	BE147147
2	31.4	31.4	749	9	AG557770 Mus muscu
3	31.2	31.2	442	8	AZ955882 2M0222110
4	30.8	30.8	696	8	B21536 F21P23-T7 I
5	30.6	30.6	461	6	CB736914 AMGNNUC.M
6	30.6	30.6	1074	2	BE371824 601217621
7	30.4	30.4	917	9	AG069018 Pan trogl
8	30	30	725	7	CK016754 AGENCOURT
9	29.8	29.8	528	8	AQ108017 CIT-HSP-2
10	29.6	29.6	578	7	CN553623 tae26g08
11	29.6	29.6	617	4	EG838990 Gm01_01e0
12	29.6	29.6	626	7	CN623811 tae59d12
13	29.6	29.6	717	7	CN553289 tae26g08
14	29.6	29.6	744	7	CN226782 RJB004H01
15	29.6	29.6	745	7	CO421969 GGEZHT100
16	29.6	29.6	852	5	BU209960 603949596
17	29.6	29.6	877	5	BU399017 603534483
18	29.2	29.2	254	4	BM401540 JH3C07R S
19	29.2	29.2	343	4	BG630802 CC-esf1cl
20	29.2	29.2	762	7	CF446768 EST683113
21	29	29	715	8	BZ944697 CH240_121
22	28.8	28.8	339	7	CN464219 7866.I AF
23	28.8	28.8	532	6	CA249294 SCSBFL110
24	28.8	28.8	605	9	CE114143 tigr-gss-

25	28.8	28.8	739	8	CC115139
26	28.6	28.6	472	8	AQ955733
27	28.6	28.6	513	5	BP051024
28	28.6	28.6	559	6	CB094897
29	28.6	28.6	561	8	BZ891642
30	28.4	28.4	580	8	BZ201271
31	28.4	28.4	607	6	CA892404
32	28.4	28.4	629	6	CA890800
33	28.4	28.4	629	6	CA892977
34	28.4	28.4	748	9	AG525698
35	28.4	28.4	756	4	BG393101
36	28.4	28.4	818	5	BU708057
37	28.4	28.4	832	9	CC491386
38	28.4	28.4	835	6	CB686108
39	28.2	28.2	489	5	BW594105
40	28.2	28.2	490	5	BW545837
41	28.2	28.2	496	5	BQ531573
42	28.2	28.2	521	5	BQ531501
43	28.2	28.2	521	7	CF791691
44	28.2	28.2	559	4	BG649176
45	28.2	28.2	598	2	BE363373

ALIGNMENTS

RESULT 1  
BE147147  
LOCUS PM2-HT0224-291099-002-c09 HT0224 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BE147147 642 bp mRNA linear EST 21-JUN-2000  
ACCESSION BE147147.1 GI:8609871  
VERSION BE147147.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 642)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=PM2-HT0224-291099-002-c09&t3=1999-10-29&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 17  
High quality sequence stop: 415.

FEATURES  
source  
1..642  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0224"  
/note="Organ: head neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products



```

QY 7 GGAGGAUUAUCCGAGCAUAGUCGACUCCGUCUCCUUAACACAGAUUAUAAUUGSU 66
DB 272 GCAGGCTTTCCATCCAGCATCTGATCTCCAGGCTTCATTTGATATAAATCTAAACAATTGAG 331
QY 67 UUVAGCAUAGCCUUAUGGACGACGAAGCUUCU 98
DB 332 TTTTGTGTTTGTGTTTTCGAGACCAGGTTTCT 363

RESULT 4
B21536 696 bp DNA linear GSS 16-SEP-1997
LOCUS F21P23-T7 IGF Arabidopsis thaliana genomic clone F21P23, genomic
DEFINITION survey sequence.
ACCESSION B21536
VERSION B21536.1 GI:2396590
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 696)
AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F21P23-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 129
High quality sequence stop: 143.
FEATURES
source
1..696
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F21P23"
/sex="hermaphrodite"
/clone_lib="IGF"
/notes="Vector: BelOBACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 30.8%; Score 30.8; DB 8; Length 696;
Best Local Similarity 43.9%; Pred. No. 14;
Matches 29; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 9 AGGAUUAUCCGAGCAUAGUCGACUCCGUCUCCUUAACACAGAUUAUAAUUGSU 68
DB 306 AGAAATCATCGAGTCAAGTTAGAAAGGTTGTTCTTGACACCACATATAGATGTGTC 365
QY 69 UAGCAU 74
DB 366 TACCAT 371

RESULT 5
CB736914/c
LOCUS CB736914
DEFINITION AMGNNUC:MRBE4-00015-C9-A rat brain E15 (10375) Rattus norvegicus
ACCESSION CB736914
VERSION CB736914.1 GI:29804127.

QY 7 GGAGGAUUAUCCGAGCAUAGUCGACUCCGUCUCCUUAACACAGAUUAUAAUUGSU 66
DB 272 GCAGGCTTTCCATCCAGCATCTGATCTCCAGGCTTCATTTGATATAAATCTAAACAATTGAG 331
QY 67 UUVAGCAUAGCCUUAUGGACGACGAAGCUUCU 98
DB 332 TTTTGTGTTTGTGTTTTCGAGACCAGGTTTCT 363

RESULT 4
B21536 696 bp DNA linear GSS 16-SEP-1997
LOCUS F21P23-T7 IGF Arabidopsis thaliana genomic clone F21P23, genomic
DEFINITION survey sequence.
ACCESSION B21536
VERSION B21536.1 GI:2396590
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 696)
AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F21P23-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 129
High quality sequence stop: 143.
FEATURES
source
1..696
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F21P23"
/sex="hermaphrodite"
/clone_lib="IGF"
/notes="Vector: BelOBACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 30.8%; Score 30.8; DB 8; Length 696;
Best Local Similarity 43.9%; Pred. No. 14;
Matches 29; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 9 AGGAUUAUCCGAGCAUAGUCGACUCCGUCUCCUUAACACAGAUUAUAAUUGSU 68
DB 306 AGAAATCATCGAGTCAAGTTAGAAAGGTTGTTCTTGACACCACATATAGATGTGTC 365
QY 69 UAGCAU 74
DB 366 TACCAT 371

RESULT 5
CB736914/c
LOCUS CB736914
DEFINITION AMGNNUC:MRBE4-00015-C9-A rat brain E15 (10375) Rattus norvegicus
ACCESSION CB736914
VERSION CB736914.1 GI:29804127.

```

```

KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 461)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00015 row: c column: 9.
FEATURES
source
1..461
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="mrbe4-00015-c9"
/tissue_type="brain E15"
/clone_lib="rat brain E15 (10375)"
/notes="Vector: pBCB; Site_1: BstXI; Site_2: NotI; rat
brain E15"

ORIGIN
Query Match 30.6%; Score 30.6; DB 6; Length 461;
Best Local Similarity 36.7%; Pred. No. 16;
Matches 29; Conservative 19; Mismatches 31; Indels 0; Gaps 0;

QY 8 GAGGAUUAUCCGAGCAUAGUCGACUCCGUCUCCUUAACACAGAUUAUAAUUGSU 67
DB 180 GAGGACTTCATCAGTAGATGTGCGATAAATTATAGTTCTTGACAGTTATGGGTGAGTT 121
QY 68 UUAGCAUAGCCUUAUGCGA 86
DB 120 TGAGCAGTTGCATTAGCAA 102

RESULT 6
B21536 1074 bp mRNA linear EST 21-JUL-2000
LOCUS 601217621F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3586861 5',
DEFINITION mRNA sequence.
ACCESSION BE371824
VERSION BE371824.1 GI:9317291
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1074)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LLAM8748 row: j column: 06
High quality sequence stop: 380.
FEATURES
source
1..1074
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

```







Laboratory of Animal Biotechnology, Dep. of Animal Production  
ESALQ - University of Sao Paulo  
Av. Padua Dias, 11, Piracicaba, SP, 13418-900, Brazil  
Tel: 55 19 3429 4434  
Fax: 55 19 3429 4285  
Email: ccsilva@esalq.usp.br and llcoutin@esalq.usp.br

PCR Primers:  
BACKWARD: T7.

FEATURES  
source Location/Qualifiers  
1..745  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="GGEZHT1005A10"  
/tissue\_type="pituitary and hypothalamus"  
/dev\_stage="21 days old"  
/lab\_host="DH10B"  
/clone\_lib="Htl"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This cDNA library was constructed with the Superscript Plasmid System with Gateway Technology kit (Invitrogen), following manufacture's protocols. Plasmid DNA was purified using a modified alkaline lysis method. Sequencing reactions were conducted using the kit Big Dye Terminator Cycle Sequencing Ready Reaction (Applied Biosystems) according to the manufacturer's recommendations. Clones were sequenced by the 5' end with T7 primer. Sequencing reactions were analyzed on ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The quality and clustering of the ESTs were analyzed using the softwares Phred/Cap3. Only EST sequences with Phred quality greater than 20 and at least 150 bp were considered for clustering."

ORIGIN  
Query Match 29.6%; Score 29.6; DB.7; Length 745;  
Best Local Similarity 55.8%; Pred. No. 37;  
Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 1 GGGAGUGGAGGAUAUCCAGCGCAUAUGUCGACUCCGUUCUCAAACC 52  
||||| : : : : : ||||| : : : : :  
Db 254 GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGTCTCCCATGCACC 305

Search completed: April 19, 2005, 16:20:16  
Job time : 3104 secs

81 UACGGACAGCAGCUU 96  
| | | | | | | | | | : :  
526 CAACGAGATCAAGTTT 541

RESULT 14  
CN226782 744 bp mRNA linear EST 09-APR-2004

LOCUS RJB004H01.ab1 RJtestis Gallus gallus CDNA 5', mRNA sequence.  
DEFINITION CN226782  
ACCESSION CN226782  
VERSION GI:46330526  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 744)  
Savolainen,P., Fitzsimmons,C.J., Arvestad,L., Andersson,L. and Lundberg,J.  
EST analysis of brain and testis cDNA libraries from White Leghorn and Red Jungle Fowl  
Unpublished (2004)  
JOURNAL Contact: Peter Savolainen  
COMMENT Department of Biotechnology  
Royal Institute of Technology, KTH  
SE-106 91 Stockholm, SWEDEN  
Tel.: +46 (0)8 5537 8481  
Fax: +46 (0)8 5537 8335  
Email: Peter.Savolainen@biotech.kth.se  
Seq primer: MJ3 reverse primer.  
Location/Qualifiers  
1..744  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Red junglefowl"  
/db\_xref="taxon:9031"  
/sex="male"  
/lab\_host="ElectroMAX DH10B (Invitrogen)"  
/clone\_lib="RJtestis"  
/note="Organ: testis; Vector: pSPORT-1; Site 1: Hind III; Site 2: EcoRI; The cDNA libraries were created with the Superscript Plasmid System (Invitrogen)."

ORIGIN  
Query Match 29.6%; Score 29.6; DB 7; Length 744;  
Best Local Similarity 55.8%; Pred. No. 37;  
Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 1 GGGAGUGGAGGAUAUCCAGCGCAUAUGUCGACUCCGUUCUCAAACC 52  
||||| : : : : : ||||| : : : : :  
Db 491 GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGTCTCCCATGCACC 542

RESULT 15  
CO421969 745 bp mRNA linear EST 02-JUL-2004

LOCUS GGEZHT1005A10.g HT1 Gallus gallus CDNA clone GGEZHT1005A10, mRNA sequence.  
DEFINITION CO421969  
ACCESSION CO421969  
VERSION GI:49638217  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 745)  
Silva,C.S., Jorge,E.C., Patricio,M., Ledur,M.C. and Coutinho,L.L.  
Discovery of new genes expressed in the chicken pituitary and hypothalamus  
Unpublished (2004)  
JOURNAL Contact: Clarissa S. Silva  
COMMENT

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2005, 13:28:09 ; Search time 130 Seconds  
(without alignments)  
1258.674 Million cell updates/sec

Title: US-09-529-397C-25  
Perfect score: 100  
Sequence: 1 gggagugaggaaucaugc.....uagcgacgaagcuucgc 100

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	30.4	30.4	366	US-09-328-352-608	Sequence 608, Appl
2	29.8	29.8	4285	US-09-410-464-1	Sequence 1, Appli
3	27.4	27.4	405	US-09-270-767-6441	Sequence 6441, Ap
4	27.4	27.4	405	US-09-270-767-21723	Sequence 21723, A
C 5	27.4	27.4	3791	US-08-675-773B-3	Sequence 3, Appli
C 6	27.4	27.4	260286	US-09-949-016-17037	Sequence 17037, A
C 7	27.4	27.4	260293	US-09-949-016-12106	Sequence 12106, A
C 8	27.0	27.0	208	US-09-513-999C-21500	Sequence 21500, A
C 9	26.8	26.8	415	US-09-513-999C-26061	Sequence 26061, A
10	26	26	6944	US-09-949-016-3637	Sequence 3637, Ap
C 11	26	26	106380	US-09-949-016-17553	Sequence 17553, A
C 12	26	26	150597	US-09-949-016-15379	Sequence 15379, A
13	25.8	25.8	798	US-08-956-171E-162	Sequence 162, App
14	25.8	25.8	798	US-08-781-986A-162	Sequence 162, App
15	25.8	25.8	77851	US-09-949-016-12508	Sequence 12508, A
16	25.8	25.8	77867	US-09-949-016-13211	Sequence 13211, A
17	25.8	25.8	77867	US-09-949-016-13212	Sequence 13212, A
18	25.8	25.8	77940	US-09-949-016-12509	Sequence 12509, A
19	25.6	25.6	372	US-08-956-171E-798	Sequence 798, App
20	25.6	25.6	372	US-08-781-986A-798	Sequence 798, App
21	25.6	25.6	27702	US-09-949-016-11795	Sequence 11795, A
22	25.4	25.4	3711	US-09-902-540-7049	Sequence 7049, Ap
C 23	25.4	25.4	13624	US-09-902-540-1053	Sequence 1053, Ap
C 24	25.4	25.4	50229	US-09-949-016-15499	Sequence 15499, A
C 25	25.4	25.4	276687	US-09-949-016-13840	Sequence 13840, A
26	25.2	25.2	355	US-09-513-999C-8830	Sequence 8830, Ap
27	25.2	25.2	601	US-09-949-016-190077	Sequence 190077,

C 28	25.2	25.2	18572	4	US-09-949-016-17183	Sequence 17183, A
29	25.2	25.2	57178	4	US-09-949-016-12838	Sequence 12838, A
30	25.2	25.2	57178	4	US-09-949-016-17458	Sequence 17458, A
C 31	25.2	25.2	786431	4	US-09-751-389-3	Sequence 3, Appli
32	25	25	246	4	US-09-513-999C-35958	Sequence 35958, A
C 33	25	25	960	4	US-09-328-352-1916	Sequence 1916, Ap
34	25	25	47781	4	US-09-949-016-16492	Sequence 16492, A
35	25	25	47781	4	US-09-949-016-16493	Sequence 16493, A
36	25	25	47781	4	US-09-949-016-16494	Sequence 16494, A
37	25	25	1830121	4	US-09-557-884-1	Sequence 1, Appli
38	25	25	1830121	4	US-09-643-990A-1	Sequence 1, Appli
39	24.8	24.8	577	4	US-09-706-228-12	Sequence 12, Appl
40	24.8	24.8	2775	4	US-09-248-796A-4695	Sequence 4695, Ap
41	24.8	24.8	3786	3	US-08-975-762-42	Sequence 42, Appl
42	24.8	24.8	3786	3	US-09-295-028-42	Sequence 42, Appl
43	24.8	24.8	3786	3	US-09-106-582-42	Sequence 42, Appl
44	24.8	24.8	3786	4	US-09-159-469-42	Sequence 42, Appl
45	24.8	24.8	3786	4	US-09-693-542-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1  
US-09-328-352-608/c  
; Sequence 608, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 608  
; LENGTH: 366  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-608

Query Match 30.4%; Score 30.4; DB 4; Length 366;  
Best Local Similarity 39.8%; Pred. No. 0.2;  
Matches 35; Conservative 17; Mismatches 36; Indels 0; Gaps 0;

Qy	1	GGGAGUGGAGGAAUUAUCGAGGCAUUAUCGACUCCGUCUUAACCAAGTAATAG 42
Db	101	GGGTGTTTATGATATAGCTTCTATATGTCACCAAGTAATGTCACCAAGTAATAG 42
Qy	61	AUUGGUUUUAGCAUUGCCUUAAGCGACA 88
Db	41	TCGTGTTTATGCGTATATCTTAGATTCA 14

RESULT 2  
US-09-410-464-1  
; Sequence 1, Application US/09410464  
; Patent No. 6395892  
; GENERAL INFORMATION:  
; APPLICANT: Strauss et al.  
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in  
; FILE OF INVENTION: poplar and other plant species.  
; FILE REFERENCE: 53375  
; CURRENT APPLICATION NUMBER: US/09/410,464  
; CURRENT FILING DATE: 1999-10-01  
; EARLIER APPLICATION NUMBER: 09/287,700  
; EARLIER FILING DATE: 1999-04-06  
; EARLIER APPLICATION NUMBER: 60/080,851  
; EARLIER FILING DATE: 1998-04-06  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4285

```
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
US-09-410-464-1

Query Match      29.8%; Score 29.8; DB 3; Length 4285;
Best Local Similarity 33.3%; Pred. No. 0.77;
Matches 27; Conservative 22; Mismatches 32; Indels 0; Gaps 0;

QY 8 GAGGAUUAUCGAGGCAUAUGCAGCUCGUCUCCUUAACACCAAGUUAUAAUUGGU 67
Db 1193 GATTAACTCTTAACCTTCCTTCCTTTTATGCTTCAATCCATCTATTATGTT 1252

QY 68 UVAGCAUUGCCUUCAGCACA 88
Db 1253 TTATGATTTTCTTAGATACA 1273

RESULT 3
US-09-270-767-6441
; Sequence 6441, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6441
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6441

Query Match      27.4%; Score 27.4; DB 4; Length 405;
Best Local Similarity 35.5%; Pred. No. 2.7;
Matches 33; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 7 GGAGGAUUAUCGAGGCAUAUGCAGCUCGUCUCCUUAACACCAAGUUAUAAUUGGU 66
Db 227 GGTGTAATCCATTTCGTTCAATTTTCGCTCTCTCAACTCAACCTCGTAAATATCCAT 286

QY 67 UUAUAGCAUUGCCUUAUGCAGCAGCAAGCUUCUG 99
Db 287 TTTTAAAGTCCTGACCAACCAACCGAGATTCTG 319

RESULT 4
US-09-270-767-21723
; Sequence 21723, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21723
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-21723

Query Match      27.4%; Score 27.4; DB 4; Length 405;
Best Local Similarity 35.5%; Pred. No. 2.7;
Matches 33; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 7 GGAGGAUUAUCGAGGCAUAUGCAGCUCGUCUCCUUAACACCAAGUUAUAAUUGGU 66
Db 227 GGTGTAATCCATTTCGTTCAATTTTCGCTCTCTCAACTCAACCTCGTAAATATCCAT 286

QY 67 UUAUAGCAUUGCCUUAUGCAGCAGCAAGCUUCUG 99
Db 287 TTTTAAAGTCCTGACCAACCAACCGAGATTCTG 319
```

```
QY 67 UUAUAGCAUUGCCUUAUGCAGCAGCAAGCUUCUG 99
Db 287 TTTTAAAGTCCTGACCAACCAACCGAGATTCTG 319

RESULT 5
US-08-675-773B-3/c
; Sequence 3, Application US/08675773B
; Patent No. 6166288
; GENERAL INFORMATION:
; APPLICANT: DIAMOND, LISA E
; APPLICANT: LOGAN, JOHN S
; APPLICANT: BYRNE, GUERARD W
; APPLICANT: SHARMA, AJAY
; TITLE OF INVENTION: METHOD OF PRODUCING TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 SEVENTH STREET, N.W., SUITE 300
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,773B
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: DIAMOND=IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-675-773B-3

Query Match      27.4%; Score 27.4; DB 3; Length 3791;
Best Local Similarity 41.6%; Pred. No. 5.7;
Matches 32; Conservative 14; Mismatches 31; Indels 0; Gaps 0;

QY 19 CGAGCAUAUUGCAGCUCGUCUCCUUAACACCAAGUUAUAAUUGGUUAGCAUAGC 78
Db 624 CCAGATTTCAGCAGACCTGCTCTCCCTCAACCCCTCTTATTAGATGTTGTGAGCATTTTC 565

QY 79 CUUAGCGCAGCAAGCU 95
Db 564 CATGGGACTTGATGCT 548

RESULT 6
US-09-949-016-17037/c
; Sequence 17037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```





Search completed: April 19, 2005, 13:30:36  
Job time : 134 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2005, 13:28:09 ; Search time 1693 Seconds  
(without alignments)  
2862.092 Million cell updates/sec

Title: US-09-529-397C-25  
Perfect score: 100  
Sequence: 1 gggaguggaggaaucaucg.....uagcgacgaagcuuucg 100

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_ste.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.2	32.2	199415	10 AC084053	AC084053 Mus muscu
C 2	31.2	31.2	201197	10 AC125279	AC125279 Mus muscu
C 3	30.8	30.8	187606	2 AC110847	AC110847 Rattus no
C 4	30.8	30.8	221981	2 AC103165	AC103165 Rattus no
C 5	30.8	30.8	231814	2 AC114165	AC114165 Rattus no
C 6	30.8	30.8	233823	2 AC110697	AC110697 Rattus no
C 7	30.8	30.8	238172	2 AC099144	AC099144 Rattus no
C 8	30.8	30.8	250740	2 AC129051	AC129051 Rattus no
C 9	30.6	30.6	748	6 CQ451733	CQ451733 Sequence
C 10	30.6	30.6	15776	6 CQ363795	CQ363795 Sequence
C 11	30.6	30.6	110000	1 AE017283_15	Continuation (16 o
C 12	30.4	30.4	217	4 AF522912	AF522912 Tachyglos
C 13	30.4	30.4	217	4 AF522914	AF522914 Tachyglos
C 14	30.4	30.4	217	4 AF522916	AF522916 Tachyglos
C 15	30.4	30.4	366	6 AR318058	AR318058 Sequence
C 16	30.4	30.4	3185	4 AV194920	AV194920 Tachyglos
C 17	30.4	30.4	46077	2 AC137289	AC137289 Rattus no
C 18	30.4	30.4	58143	2 AC137177	AC137177 Rattus no
C 19	30.4	30.4	207260	2 AC134220	AC134220 Rattus no

20	30.4	30.4	213033	2	AC131482	AC131482 Rattus no
C 21	30.4	30.4	218157	2	AC139588	AC139588 Rattus no
C 22	30.4	30.4	272400	2	AC110146	AC110146 Rattus no
C 23	30.4	30.4	283299	2	AC112576	AC112576 Rattus no
C 24	30.2	30.2	191674	4	CR536601	CR536601 Platypus
C 25	30.2	30.2	229522	2	AC109911	AC109911 Rattus no
C 26	30.2	30.2	235937	2	AC131223	AC131223 Rattus no
C 27	30.2	30.2	240755	2	AC095368	AC095368 Rattus no
C 28	30.2	30.2	240755	2	AC133102	AC133102 Mus muscu
C 29	30.0	30.0	207191	2	AR372457	AR372457 Sequence
C 30	29.8	29.8	4285	6	AF057708	AF057708 Populus b
C 31	29.8	29.8	184147	9	AC103805	AC103805 Homo sapi
C 32	29.8	29.8	202287	2	AC145887	AC145887 Pan trogl
C 33	29.8	29.8	215994	9	AC067941	AC067941 Homo sapi
C 34	29.8	29.8	223143	2	CR450730	CR450730 Danio rer
C 35	29.8	29.8	227025	2	AC148836	AC148836 Pan trogl
C 36	29.6	29.6	23393	2	AC018216	AC018216 Drosophil
C 37	29.6	29.6	50481	2	AC117541	AC117541 Mus muscu
C 38	29.6	29.6	79826	3	AC002443	AC002443 Drosophil
C 39	29.6	29.6	110000	2	AC108323	Continuation (4 of
C 40	29.6	29.6	110000	2	AC112373_04	Continuation (5 of
C 41	29.6	29.6	156806	3	AC093047	AC093047 Drosophil
C 42	29.6	29.6	182816	5	AC145913	AC145913 Gallus ga
C 43	29.6	29.6	192132	3	AC093440	AC093440 Drosophil
C 44	29.6	29.6	234627	2	AC106118	AC106118 Rattus no
C 45	29.6	29.6	237005	2	AC137343	AC137343 Rattus no

## ALIGNMENTS

RESULT 1	AC084053	199415 bp	DNA	linear	ROD 05-SEP-2002
LOCUS	Mus Musculus Strain C57BL6/J chromosome 3 BAC, RP23-158F1, complete sequence.				
DEFINITION	AC084053				
ACCESSION	AC084053				
VERSION	AC084053.6	GI:22725947			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.				
AUTHORS	High Throughput Mouse Sequencing				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 199415)				
REFERENCE	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (12-OCT-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA				
JOURNAL	3 (bases 1 to 199415)				
REFERENCE	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (29-JUL-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA				
JOURNAL	4 (bases 1 to 199415)				
REFERENCE	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (05-SEP-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA				
JOURNAL					





```

repeat_region 23882..24118
/rpt_family="Lx4"
repeat_region complement(24228..24641)
/rpt_family="Lx6"
repeat_region complement(24821..25092)
/rpt_family="Lx3"
repeat_region 25183..25208
/rpt_family="(TAA)n"
repeat_region 25210..25277
/rpt_family="(GAA)n"
repeat_region complement(25288..25661)
/rpt_family="MTC"
repeat_region complement(26114..32348)
/rpt_family="LIF"

Query Match 32.2%; Score 32.2; DB 10; Length 199415;
Best Local Similarity 49.1%; Pred. No. 8.6;
Matches 26; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 37 CGUCUCCUUAACACAGUUAUUAUGUUUAGCAUUGCCUAGCGACAG 89
Db 63221 CTTATTCTTAAACACAGTTATAAATCTTGCTAGCATATGCAATAGTGTCTG 63273

RESULT 2
AC125279/c
LOCUS AC125279 201197 bp DNA linear ROD 05-NOV-2003
DEFINITION Mus musculus BAC clone RP23-433F5 from 8, complete sequence.
ACCESSION AC125279
VERSION AC125279.1 GI:21536173
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201197)
Trani, L. and Cotton, M.
The sequence of Mus musculus BAC clone RP23-433F5
Unpublished (2001)
2 (bases 1 to 201197)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 201197)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 201197)
Wilson, R.
Direct Submission
Submitted (05-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submission@wustl.wustl.edu
----- Summary Statistics
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tatenio in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES	source	Location/Qualifiers
repeat_region	1..201197	/organism="Mus musculus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10090"
		/chromosome="8"
		/map="8"
		/clone="RP23-433F5"
		/clone_lib="RPCI-23"
repeat_region	171..557	/rpt_family="MaLR"
repeat_region	558..1352	/rpt_family="MaLR"
repeat_region	1353..1499	/rpt_family="Alu"
repeat_region	1500..1886	/rpt_family="MaLR"
repeat_region	1887..2101	/rpt_family="MaLR"
repeat_region	2103..2485	/rpt_family="L1"
repeat_region	2792..2918	/rpt_family="Alu"
repeat_region	2916..3397	/rpt_family="RMR6B"
repeat_region	5516..5712	/rpt_family="MaLR"
repeat_region	6071..6162	/rpt_family="ERVK"
repeat_region	6163..6556	/rpt_family="MaLR"
repeat_region	6557..6980	/rpt_family="ERVK"
repeat_region	7137..7220	/rpt_family="MIR"
repeat_region	7970..8262	/rpt_family="L1"
repeat_region	8734..9125	/rpt_family="L1"
repeat_region	9194..10473	/rpt_family="L1"
repeat_region	10485..10619	/rpt_family="MER1_type"
repeat_region	10573..10940	/rpt_family="ERVK"
repeat_region	10673..11051	/rpt_family="RMR17C"
repeat_region	11200..12140	/rpt_family="L1"
repeat_region	12674..12714	/rpt_family="L1"
repeat_region	12715..12907	/rpt_family="B2"
repeat_region	12908..13742	/rpt_family="L1"

```

repeat_region 14760..15436
/rpt_family="L1"
repeat_region 16028..16131
/rpt_family="MaLR"
repeat_region 17836..18263
/rpt_family="L1"
repeat_region 18352..18653
/rpt_family="ERVK"
repeat_region 18655..19078
/rpt_family="ERVK"
repeat_region 19375..19942
/rpt_family="L1"
repeat_region 21030..21117
/rpt_family="MaLR"
repeat_region 22308..22444
/rpt_family="MIR"
repeat_region 22775..22988
/rpt_family="MER1_type"
repeat_region 22991..23088
/rpt_family="L1"
repeat_region 23274..23681
/rpt_family="MaLR"
repeat_region 24784..24971
/rpt_family="B2"
repeat_region 25314..25486
/rpt_family="MaLR"
repeat_region 26145..26223
/rpt_family="RAL_RN"
repeat_region 26473..26619
/rpt_family="Alu"
repeat_region 26817..27001
/rpt_family="B2"
tRNA complement(26924..26996)
/product="tRNA-Ser"
/notes="Likely pseudogene (HMM Sc=39.59 / Sec struct
Sc=-11.21)"
repeat_region 27016..27079
/rpt_family="Alu"
repeat_region 28465..28604
/rpt_family="Alu"
repeat_region 28818..29133
/rpt_family="MaLR"
repeat_region 29251..29389
/rpt_family="MIR"
repeat_region 29545..29687
/rpt_family="MaLR"
repeat_region 30593..30737
/rpt_family="L2"
repeat_region 32073..32437
/rpt_family="MaLR"
repeat_region 33137..33276
/rpt_family="Alu"
repeat_region 33402..33547
/rpt_family="Alu"
repeat_region 35207..35290
/rpt_family="MIR"
repeat_region 35653..35830
/rpt_family="MaLR"
repeat_region 35831..36039
/rpt_family="B2"
repeat_region 36046..36232
/rpt_family="MaLR"
repeat_region 36324..36549
/rpt_family="B4"
repeat_region 36709..36788
/rpt_family="B4"
repeat_region 36795..36891
/rpt_family="ERVL"
repeat_region 36862..37099
/rpt_family="MaLR"
repeat_region 39693..39806
/rpt_family="MER1_type"
repeat_region 40903..41102

```

```

/rpt_family="MER2_type"
41131..41296
/rpt_family="MER2_type"
41519..41626
/rpt_family="L2"
42585..42730
/rpt_family="Alu"

Query Match 31.2%; Score 31.2; DB 10; Length 201197;
Best Local Similarity 32.6%; Pred. No. 19;
Matches 30; Conservative 24; Mismatches 38; Indels 0; Gaps 0;

QY 7 GGAGGAUUCUACGAGGCAUUGUGACUCGCUUUCUUAACCAACGUAUUAUUGU 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2984 GCAGGCTTTTCATCCAGCATCTGATTCCTGAGTCTTCATTGATAAATCTACATTGAG 2925

QY 67 UUUAGCAUAGCCUAGGACGACGACGAAGCUUCU 98
::: : : : : : : : : : : : : : : : : : : : : :
Db 2924 TTTTGTGTTTGTGTTTTCGAGACCAGGTTTCT 2893

RESULT 3
AC110847
LOCUS
DEFINITION
AC110847 187606 bp DNA linear HTG 15-NOV-2002
Rattus norvegicus clone CH230-42M10, *** SEQUENCING IN PROGRESS
**, 13 unordered pieces.
AC110847 GI:25006792
AC110847.6 HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 187606)
Munzy,D.Marie., Metzker,M.Iee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Liu,J.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackeleme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,N., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

```

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 187606)  
Worley, K.C.  
Direct Submission  
Submitted (16-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 187606)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23321762.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GRXS  
Center clone name: CH230-42M10  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 118671 bases at least Q40  
Consensus quality: 124131 bases at least Q30  
Consensus quality: 127839 bases at least Q20  
Estimated insert size: 122165; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 31167: contig of 31167 bp in length  
\* 31168 31267: gap of unknown length  
\* 77383 77383: contig of 46116 bp in length  
\* 77384 77383: gap of unknown length  
\* 77484 168911: contig of 91428 bp in length  
\* 168912 168911: gap of unknown length  
\* 169012 170421: contig of 1410 bp in length  
\* 170422 170521: gap of unknown length  
\* 170522 172029: contig of 1508 bp in length

\* 172030 172129: gap of unknown length  
\* 172130 173707: contig of 1578 bp in length  
\* 173708 173807: gap of unknown length  
\* 173808 175104: contig of 1297 bp in length  
\* 175105 175204: gap of unknown length  
\* 175205 176766: contig of 1562 bp in length  
\* 176767 176866: gap of unknown length  
\* 176867 178988: contig of 2122 bp in length  
\* 178989 179088: gap of unknown length  
\* 179089 180430: contig of 1342 bp in length  
\* 180431 180530: gap of unknown length  
\* 180531 182073: contig of 1543 bp in length  
\* 182074 182173: gap of unknown length  
\* 182174 184491: contig of 2318 bp in length  
\* 184492 184591: gap of unknown length  
\* 184592 187606: contig of 3015 bp in length.

FEATURES  
Location/Qualifiers  
source  
1..187606  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-42M10"  
misc\_feature  
13985..15402  
/note="wgs\_contig"  
misc\_feature  
28318..29829  
/note="wgs\_contig"  
misc\_feature  
77484..79244  
/note="wgs\_contig"  
misc\_feature  
95087..96968  
/note="wgs\_contig"  
misc\_feature  
131603..132754  
/note="wgs\_contig"  
misc\_feature  
145131..150091  
/note="wgs\_contig"  
misc\_feature  
167411..168911  
/note="wgs\_contig"

ORIGIN  
Query Match 30.8%; Score 30.8; DB 2; Length 187606;  
Best Local Similarity 34.7%; Pred. No. 26;  
Matches 34; Conservative 22; Mismatches 42; Indels 0; Gaps 0;  
OY 2 GGAGUGGAGGAUAUCAUGGAGCAUAGUGGAGCUCGUCUCCUACCAACGAGUUAUA 61  
Db 164680 GTAGACGAGATATCATTTAGGAGTTATTATTAATCACTGTTTCTTAAGATCATTTAT 164739  
OY 62 UUGGUUAGCAUAGCUCUAGCGACGACGACGACUUCUG 99  
Db 164740 TTAGTTTTCCTAGTCTCTGGGCCATCAAGTTTCAG 164777

RESULT 4  
AC103165/c  
LOCUS AC103165 221981 bp DNA linear HTG 13-MAY-2003  
DEFINITION Rattus norvegicus clone CH230-194N19, WORKING DRAFT SEQUENCE.  
ACCESSION AC103165  
VERSION AC103165.6 GI:30578614  
HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 221981)  
AUTHORS Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, N., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, R., Chen, G., Chen, Y., Chen, Y., Chen, J., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,





Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 233823)  
 Muzny,D,Marie, Metzker,M, Lee, Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguliano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensheva,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawlin,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,  
 Nwaokelumen,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shateman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steidle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tincey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 233823)  
 Worley,K.C.  
 Direct Submission  
 Submitted (15-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 233823)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (23-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 23, 2002 this sequence version replaced gi:23267238.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRXR  
 Center clone name: CH230-43H7  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 212738 bases at least Q40  
 Consensus quality: 215401 bases at least Q30  
 Consensus quality: 217067 bases at least Q20  
 Estimated insert size: 219523; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 233823: contig of 233823 bp in length.  
 -----

## FEATURES

source  
 1..233823  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-43H7"  
 misc\_feature  
 1..1276  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 misc\_feature  
 1327..2821  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 misc\_feature  
 7344..8215  
 /note="clone\_boundary  
 clone\_end:T7  
 site:ECORI  
 end\_sequence:BH264888"  
 misc\_feature  
 85851..88514  
 /note="wgs\_contig"  
 complement(232161..232332)  
 /note="clone\_boundary  
 clone\_end:Sp6  
 site:ECORI  
 end\_sequence:BH264890"  
 misc\_feature  
 complement(232161..232332)  
 /note="clone\_boundary  
 clone\_end:Sp6  
 site:ECORI  
 end\_sequence:BH264890"

## ORIGIN

Query Match 30.8%; Score 30.8; DB 2; Length 233823;  
 Best Local Similarity 34.7%; Pred. No. 26;  
 Matches 34; Conservative 22; Mismatches 42; Indels 0; Gaps 0;  
 QY 2 GGAGUGGAGGAUAUUGCAGGCAUAUGCAGCUCCUUCUUAACACGUAUAAA 61  
 DB 230212 GTAGAGCAGATATCATTTAGGAGTTATTTAATCACTGTTTTTCTAAAGTCATTATTAT 230271  
 QY 62 UUGGUUUGCAUUGCCUAGCGCAGCAAGCUUCUG 99  
 DB 230272 TTAGTTTTCCTAGTCTCTGGGCCATCAAGTTTCAG 230309

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23264509.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-68F7  
Center Clone name: CH230-68F7  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 220106 bases at least Q40  
Consensus quality: 222963 bases at least Q30  
Consensus quality: 224727 bases at least Q20  
Estimated insert size: 230633; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html.)  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is, believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 238172: contig of 238172 bp in length.

Location/Qualifiers  
1. 238172  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-68F7"  
1. 1721  
/note="wgs end\_extension  
clone\_end:Sp6"  
8345..18845  
/note="clone\_boundary  
clone\_end:Sp6  
site:ECORI  
end\_sequence:BH351992"  
234341..234906  
/note="clone\_boundary  
clone\_end:T7  
site:ECORI  
end\_sequence:BH351956"  
236172..238172  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

FEATURES  
source

ORIGIN  
Query Match 30.8%; Score 30.8; DB 2; Length 238172;  
Best Local Similarity 43.1%; Pred. No. 26;  
Matches 25; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

QC 23 GCAUAGUCGACUCGUCUCUCUACACCAAGUUAUAAUUGUUUAGCAUAGCCU 80







Query Match 30.6%; Score 30.6; DB 6; Length 15776;  
Best Local Similarity 46.8%; Pred. No. 25;  
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

QY 18 UCGAGGCAUAUGUCACUCCGUUUCCUCAAACACGUAUAUAAUUGGUUUUAGCAUAUG 77  
          :  
Db 8427 TTGGGTGCAGCTTCAGCCGACGGTCTTCAGACCGTTGATGAACGGTTGTAGGT CATG 8486

QY 78 CCUAGCGACAGCAAGC 94  
      || : ||| |||||  
Db 8487 CCCTGGGCACGGGAAGC 8503

RESULT 11  
AE017283\_15  
WPCOMMENT

Sequence split into 26 fragments LOCUS AE017283 Accession AE017283

Fragment Name	Begin	End
AE017283_00	1	110000
AE017283_01	100001	210000
AE017283_02	200001	310000
AE017283_03	300001	410000
AE017283_04	400001	510000
AE017283_05	500001	610000
AE017283_06	600001	710000
AE017283_07	700001	810000
AE017283_08	800001	910000
AE017283_09	900001	1010000
AE017283_10	1000001	1110000
AE017283_11	1100001	1210000
AE017283_12	1200001	1310000
AE017283_13	1300001	1410000
AE017283_14	1400001	1510000
AE017283_15	1500001	1610000
AE017283_16	1600001	1710000
AE017283_17	1700001	1810000
AE017283_18	1800001	1910000
AE017283_19	1900001	2010000
AE017283_20	2000001	2110000
AE017283_21	2100001	2210000
AE017283_22	2200001	2310000
AE017283_23	2300001	2410000
AE017283_24	2400001	2510000
AE017283_25	2500001	2560265

Continuation (16 of 26) of AE017283 from base 1500001. (AE017283 Propionibacterium acnes)

Query Match 30.6%; Score 30.6; DB 1; Length 110000;  
Best Local Similarity 46.8%; Pred. No. 29;  
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

QY 18 UCGAGGCAUAUGUCACUCCGUUUCCUCAAACACGUAUAUAAUUGGUUUUAGCAUAUG 77  
          :  
Db 34666 TTGGGTGCAGCTTCAGCCGACGGTCTTCAGACCGTTGATGAACGGTTGTAGGT CATG 34725

QY 78 CCUAGCGACAGCAAGC 94  
      || : ||| |||||  
Db 34726 CCCTGGGCACGGGAAGC 34742

RESULT 12  
AF522912/c  
LOCUS  
DEFINITION Tachygllossus aculeatus clone Tag7 type I interferon gene, partial cdb.  
ACCESSION AF522912  
VERSION AF522912.1 GI:27451581  
KEYWORDS  
SOURCE Tachygllossus aculeatus (Australian echidna)  
ORGANISM Tachygllossus aculeatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Monotremata; Tachygllossidae; Tachygllossus.  
REFERENCE 1 (bases 1 to 217)



Search completed: April 19, 2005, 13:59:03  
Job time : 1700 secs

**THIS PAGE BLANK (USPTO)**